NPS55ZE72061A

NAVAL POSTGRADUATE SCHOOL

Monterey, California

SOME ALTERNATIVES TO EXPONENTIAL SMOOTHING IN DEMAND FORECASTING

by

Peter W. Zehna

June 1972

39

UNCLASSIFIED
Security Classification

37

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Some Alternatives to Exponential Smoothing in Demand Forecasting

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Peter W. Zehna

6. REPORT DATE June 1972	7a. TOTAL NO. OF PAGES 41	7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. NAVSUP RDT&E No. 38.531.001 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) NPS55ZE72061A 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Research and Development Division Naval Supply Systems Command

13. ABSTRACT

A study devoted to a comparison of exponential smoothing with other alternatives to demand forecasting. Special attention is paid to the stock-out risks assumed whenever reorder levels are set using the various methods being compared. Models presently used by NavSup are employed in order that the results be applicable to the system in use. Simulation techniques are used for drawing comparisons. For constant mean, normal demand, it is shown that exponential smoothing does not produce as accurate results as ordinary maximum likelihood techniques. For the case of a linear mean changing with time, it is shown that the two methods are about comparable. Finally, a sequential Bayes forecasting method is defined and found to compare quite favorably with exponential smoothing. The need for additional study of Bayesian methods is established.

I

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Inventory Theory						
Exponential Smoothing						
Forecasting						
Mean Absolute Deviation						
Reorder Levels						
Bayesian Methods						
Maximum Likelihood Estimation						

II

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral A. S. Goodfellow, USN
Superintendent

M. U. Clauser
Provost

ABSTRACT

A study devoted to a comparison of exponential smoothing with other alternatives to demand forecasting. Special attention is paid to the stock-out risks assumed whenever reorder levels are set using the various methods being compared. Models presently used by NavSup are employed in order that the results be applicable to the system in use. Simulation techniques are used for drawing comparisons. For constant mean, normal demand, it is shown that exponential smoothing does not produce as accurate results as ordinary maximum likelihood techniques. For the case of a linear mean changing with time, it is shown that the two methods are about comparable. Finally, a sequential Bayes forecasting method is defined and found to compare quite favorably with exponential smoothing. The need for additional study of Bayesian methods is established.

This task was supported by the Research and Development Division, Naval Supply Systems Command.

Prepared by:

Peter W. Zehna

Peter W. Zehna

Approved by:

J. R. Borsting, Chairman
Department of Operations Research and Administrative Sciences

Released by:

Charles S. Wagner for

C. E. Menneken
Dean of Research Administration

NPS55ZE72061A

TABLE OF CONTENTS

		Page
1.	Introduction	1
2.	Notation and Summary of Previous Results	5
3.	Linear Mean Model	14
4.	A Bayes Procedure	20
5.	Concluding Remarks	32
	Bibliography	34

1. INTRODUCTION.

In two previous reports ([1] and [2]) a rather detailed examination of some of the aspects of exponential smoothing as a demand forecasting tool was presented. In particular, special attention was paid to the manner in which reorder levels are affected in a variety of forms using models presently employed by NavSup and originally generated by R. G. Brown [3].

In the case of a normal demand with constant mean and variance (high mover, low value items) the results of setting reorder levels using exponential smoothing were compared with those obtained using classical maximum likelihood techniques. Because of the intractability of the probability distributions involved using exponential smoothing, simulation techniques had to be used for comparing these methods. While such methods fail to produce absolutely conclusive results, the overwhelming evidence favoring maximum likelihood over exponential smoothing in every case examined can hardly be taken lightly. The results really were not surprising. As previously pointed out, whenever the Gauss-Markov assumptions apply, as they do in these models, almost any departure from maximum likelihood methods is doomed to be second best, at most. Yet, on the practical side, one can ask, "How bad off is second best?" and, "Is there a trade-off perhaps between optimality and some other desirable facets such as reduced computation time or perhaps ease of understanding?" Again, attempts were made to answer these questions by having NavSup personnel choose the criterion

and then draw comparisons on that criterion. For the models studied, very little beyond the intuitive appeal of weighting previous demands with the highest weight going to the most recent demand could be said for exponential smoothing.

To be more specific, previous studies focused on the case where demand in a period is normally distributed with mean μ and standard deviation σ. From period to period, such demands are independent but always with this same probability distribution. If μ and σ were known, then it would be a relatively simple matter to set a reorder level to apply period by period in order to achieve a specified stockout risk. Indeed, if X represents random demand in the period to come and a stockout risk of ρ is specified, then the reorder level should be set at $\mu + k\sigma$ where the constant k is determined from the simple relationship,

$$\rho = P(X>\mu+k\sigma) \tag{1.1}$$

Since this can be immediately translated into

$$\rho = P(\frac{X-\mu}{\sigma} > k) \tag{1.2}$$

and $\frac{X-\mu}{\sigma}$ is the standard or tabled normal random variable, it is a trivial task to match k with ρ by means of a normal table. For example, if $\rho = .05$, then $k = 1.645$, while if $\rho = .10$, then $k = 1.282$ and so on. Obviously, choosing larger and larger values

of k guards against being out of stock, but only at the expense perhaps of holding excessive stock on hand. The difference in the consequences of these two standard undesirable conditions will have to guide one's choice of ρ hence k.

The difficulty is that even if the model applies, the parameters μ and σ are rarely known. This means that they will have to be estimated and when these estimates are used to set the reorder level, there is no longer any guarantee that the specified value of ρ in (1.1) is satisfied. This is true regardless of how μ and σ are estimated and is just one of those statistical facts of life. The true risk that is faced thus depends upon the joint probability distribution of the estimators involved and may or may not depart significantly from the intended risk. Or if you prefer, the actual costs of being out of stock will eventually be observed to depart from what was supposed to be the case because of the fact that the estimated reorder level is not the theoretical one specified by (1.1).

This being the case, the precision with which μ and σ are estimated becomes an extremely important factor. And here is precisely where exponential smoothing begins to lose contests, at least in the normal models that have been examined. The numerical results in all of those cases, coupled with some theoretical results to be reported presently, indicates that exponential smoothing always seems to be more variable than classical maximum likelihood. What is worse, that variance does not improve with time, is a function of the

smoothing constant and, in that regard, can only be reduced at the expense of destroying the most compelling reason for employing it, namely, reducing the weight assigned to the most recent observation to zero.

It has been brought to the writer's attention that exponential smoothing really was never "invented" for the constant mean model in the first place. Perhaps so, but it is, nevertheless, presently used in precisely those cases and hence must stand on its own merit under scrutiny, particularly when alternatives are available that appear to do a better job for an equal amount of effort. Of even more significance, however, is the fact that exponential smoothing was found to be second best even in one case where the mean value of the demand process is allowed to change in time. These results are reported in Section 3.

Before turning to specific results, perhaps a remark or two regarding random demand would be in order. Generally speaking, if demand is truly random and the values of these random variables are used to set reorder levels, or in general estimate parameters, it is inherently part of the model that the resulting values will fluctuate in a random fashion also. There is no way around this point and usually the best we can hope for is that these random fluctuations eventually dampen about some ideal or hope-for value. First, we usually try to establish that at least these resultant processes will converge to a target value in the mean. Thus, it is desirable certainly to be able

to establish that random reorder levels will eventually converge in expected value to $\mu + k\sigma$ whatever μ and σ happen to be. But, such convergence is not enough. Unless the variance of that process goes to zero in time there is no assurance that the process is in any sense close to the required value regardless of how long the system may have been operating. It is this examination of variance properties of exponential smoothing that is notably lacking in the published literature. In this report, such considerations are included in a detailed examination of several models currently in vogue.

2. NOTATION AND SUMMARY PREVIOUS RESULTS.

Perhaps it is unfair to indict exponential smoothing as being the fundamental problem in the models tested. In a previous report [2], it was pointed out that it is a combination of exponential smoothing with the use of mean absolute deviation (MAD) as a means of estimating variability that appears to create the major difficulty. To summarize this point and report additional results, the following notation is adopted.

Let $X_0, X_1, X_2, \ldots, X_t$ be a demand record through time t. We assume for this section that these are mutually independent normal random variables each with mean μ and standard deviation σ. Following Brown [3], we let $\hat{X}_{t-1}$ denote the forecast at time $t - 1$ of the demand in the $t^{\underline{th}}$ period using exponential smoothing of the data to compute its value.

$$\hat{X}_{t-1} = \alpha \sum_{k=0}^{t-2} \beta^k X_{t-1-k} + \beta^{t-1} X_0 \qquad 0 < \alpha < 1; \quad \beta = 1 - \alpha \tag{2.1}$$

By using this basic formula, it can be shown [1] that $E[\hat{X}_{t-1}] = \mu$ for all t so that we may view (2.1) as an unbiased estimator of mean demand μ from period to period. If we then define a forecast error at time t by means of the formula

$$e_t = X_t - \hat{X}_{t-1} \tag{2.2}$$

then it follows that $E(e_t) \equiv 0$.

However, as previously remarked, the variance of any estimator must also be examined. In a previous report, we established that

$$\mathrm{Var}(\hat{X}_{t-1}) = \frac{\alpha + 2\,\beta^{2t-1}}{2 - \alpha}\,\sigma^2 \tag{2.3}$$

Asymptotically then,

$$\mathrm{Var}(\hat{X}_{t-1}) \to \frac{\alpha}{2 - \alpha}\,\sigma^2 \quad \text{as} \quad t \to \infty \tag{2.4}$$

Now this is a positive constant and it must be recognized then that, as an estimator for μ, $\hat{X}_{t-1}$ can never be more precise than this limiting variance allows. In other words, no matter how long the system has operated, the forecast will fluctuate about μ with a variance whose size depends upon the unknown variance σ^2 as well as, of course, the choice of the smoothing constant α.

The same remarks can also be made about the forecast error e_t. Although its expected value vanishes for all t, it too has a limiting variance bounded away from zero and given by the formula

$$(2.5) \qquad \sigma_e^2 = \lim_{t\to\infty} \mathrm{Var}(e_t) = \lim_{t\to\infty} \frac{2 + 2\beta^{2t-1}}{2-\alpha} \quad = \frac{2}{2-\alpha}\,\sigma^2$$

This result also allows us to write σ, an unknown parameter, in terms of the limiting standard deviation σ_e as,

$$(2.6) \qquad \sigma = \sqrt{\frac{2-\alpha}{2}}\,\sigma_e$$

The main reason for noting this relationship is to comply with the NavSup procedure for estimating σ by means of estimates of σ_e. These are in turn found by smoothed estimates of MAD. In the normal case, which is the only one we are treating, σ_e is related to MAD, Δ_e by means of the formula,

$$(2.7) \qquad \sigma_e = \sqrt{\frac{\pi}{2}}\,\Delta_e$$

Combining this with (2.6) yields,

$$(2.8) \qquad \sigma = \frac{\sqrt{\pi(2-\alpha)}}{2}\,\Delta_e$$

Exponentially smoothed estimates of Δ_e are obtained by smoothing forecast errors. By formula,

$$(2.9) \qquad \tilde{\Delta}_e = \alpha \sum_{k=0}^{t-1} \beta^k \, |e_{t-k}|$$

If this result is substituted ad hoc into (2.8) one then obtains the estimate

$$\tilde{\sigma} = \frac{\sqrt{\pi(2-\alpha)}}{2} \tilde{\Delta}_e \tag{2.10}$$

consistent with formulas established by Brown. We are then but a step away from the formula for setting a reorder level using smoothed estimates. First, the constant mean is estimated. After t periods of demand have been observed, mean demand is estimated by means of the formula,

$$\tilde{\mu} = \alpha \sum_{k=0}^{t-1} \beta^k X_{t-k} \tag{2.11}$$

The formula ignores initial conditions which are rendered ineffectual in time anyway. Since the claims for smoothing properties are asymptotic in the first place, this represents no serious modification and yields at least an asymptotic unbiasedness wherein $E(\tilde{\mu}) \to \mu$. When this estimate is combined with (2.10), a smoothed estimate of the reorder level becomes

$$\tilde{R} = \tilde{\mu} + k\,\tilde{\sigma} \tag{2.12}$$

where k is chosen to satisfy a required stock-out risk ρ as determined by (1.1).

As previously noted, however, the true risk that is achieved by using (2.12), or indeed any formula involving only estimates of

μ and σ, will depend on how well those parameters are estimated. Of special interest is the comparison of smoothed estimates with maximum likelihood methods wherein,

$$\hat{R} = \hat{\mu} + k\,\hat{\sigma} \tag{2.13}$$

with $\hat{\mu} = \frac{1}{t}\sum_{i=1}^{t} X_i$ and $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{t}(X_i-\hat{\mu})^2}{t}}$

being the ordinary maximum likelihood estimates of μ and σ. This point was the subject of some of the discussion in [2]. It was pointed out there in several ways that (2.13) was superior to (2.12) in case after case. Subsequent examinations by Ornek [4] and Coventry [5] reveal the same consistent behavior.

While all of these results continue to be based on simulations, the consistency with which exponential smoothing tends to produce more variable results than maximum likelihood cannot be ignored. Moreover, there is now some theoretical basis for this claim. Ornek has been able to establish an exact formula for the asymptotic variance of $\tilde{\Delta}_e$ which is of course a fundamental quantity used in the computation of a reorder level. The expression is complicated and is not duplicated here; details may be found in [4]. For all practical purposes approximate values with a high degree (within 10^{-10}) of accuracy were computed for various choices of the smoothing constant α. A summary appears in Table 2.1.

α	$\mathrm{Var}(\tilde{\Delta}_e/\sigma^2)$	S.D. of $\tilde{\sigma}/\sigma$
.10	.0204	.1745
.15	.0325	.2173
.20	.0461	.2553
.25	.0614	.2905
.30	.0785	.3239
.35	.0979	.3561
.40	.1196	.3876
.45	.1440	.4187
.50	.1716	.4495
.55	.2026	.4802
.80	.4267	.6341

Table 2.1. Variability of $\tilde{\Delta}_e$ and $\tilde{\sigma}$.

The table amply demonstrates how asymptotic variability increases with the choice of α but more importantly perhaps, no matter how long the system runs, the variance of $\tilde{\sigma}$ never approaches zero and is bounded away by a positive quantity. This means that estimates of σ, and hence of R, the theoretical reorder level, are doomed to fluctuate forever. Not so for maximum likelihood. It is well known that the variance of $\hat{\sigma}$ goes to zero with increasing t (as does the variance of $\hat{\mu}$ of course) so that eventually, $\hat{R}$ and R coincide for all practical purposes. Put another way, the intended risk ρ and the actual risk attained will be the same, whereas the same statement simply cannot be made about $\tilde{R}$.

All of this merely supports what was already observed in simulation results. Extending those results already established in the pilot study of [2], simulations were run for various parameter pairs and the risk levels compared at the 1,000th observation. For each of several such parameter pairs, five risk levels were chosen. Then actual risks $\tilde{\rho}$ for smoothed estimates were compared with actual risks $\hat{\rho}$ using maximum likelihood techniques. These results are reported in Table 2.2 and they pretty well speak for themselves. The attained risks, as measured by $\hat{\rho}$, are consistently nearer the target value ρ than are those determined by $\tilde{\rho}$. What this means is that even after the system has operated for a long, long time, with initial conditions and other factors stabilized, the actual risk attained when reorder levels are set using (2.12) may in any period be significantly different from the value that presumably was being attained by the choice of k.

Another way to view the greater variability involved when smoothing is used to set reorder levels over a long period of time was devised by Coventry. For this experiment parameter values of $\mu = 100$ and $\sigma = 10$ were chosen. Using a risk level of .05, the theoretical reorder level would be 116.45. Demands were generated for 1,000 periods and the reorder level using $\hat{R}$ and $\tilde{R}$ was checked at the 1000th period. This experiment was then replicated 100 times and the various values of $\hat{R}$ and $\tilde{R}$ were checked and plotted against the theoretical reorder level. The results are displayed in Figure 2.1

Parameters	ρ = .01		ρ = .05		ρ = .11		ρ = .25		ρ = .50	
(μ,σ)	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$
(10,1)	.044	.013	.119	.061	.136	.126	.333	.276	.537	.529
(20,2)	.092	.011	.235	.053	.364	.113	.556	.251	.772	.427
(30,3)	.058	.012	.156	.055	.262	.116	.415	.253	.635	.494
(40,4)	.002	.007	.018	.041	.050	.097	.145	.237	.369	.502
(50,5)	.000	.009	.002	.046	.014	.103	.086	.240	.371	.493
(110,10)	.000	.009	.003	.048	.020	.108	.125	.249	.489	.507
(200,20)	.005	.009	.026	.046	.061	.103	.152	.238	.350	.489
(300,30)	.003	.010	.028	.049	.078	.108	.220	.246	.511	.495
(400,40)	.021	.008	.057	.044	.096	.101	.173	.238	.312	.492
(500,50)	.005	.015	.032	.063	.075	.127	.189	.264	.422	.499
(600,60)	.051	.009	.165	.048	.287	.106	.490	.244	.783	.497
(700,70)	.015	.011	.070	.052	.150	.111	.322	.247	.598	.490
(800,80)	.068	.012	.157	.055	.239	.116	.374	.253	.563	.498
(900,90)	.028	.010	.080	.048	.137	.107	.248	.245	.430	.497
(1000,100)	.001	.009	.014	.046	.045	.102	.148	.232	.405	.475

Table 2.2. Comparison of Risks After 1,000 Periods

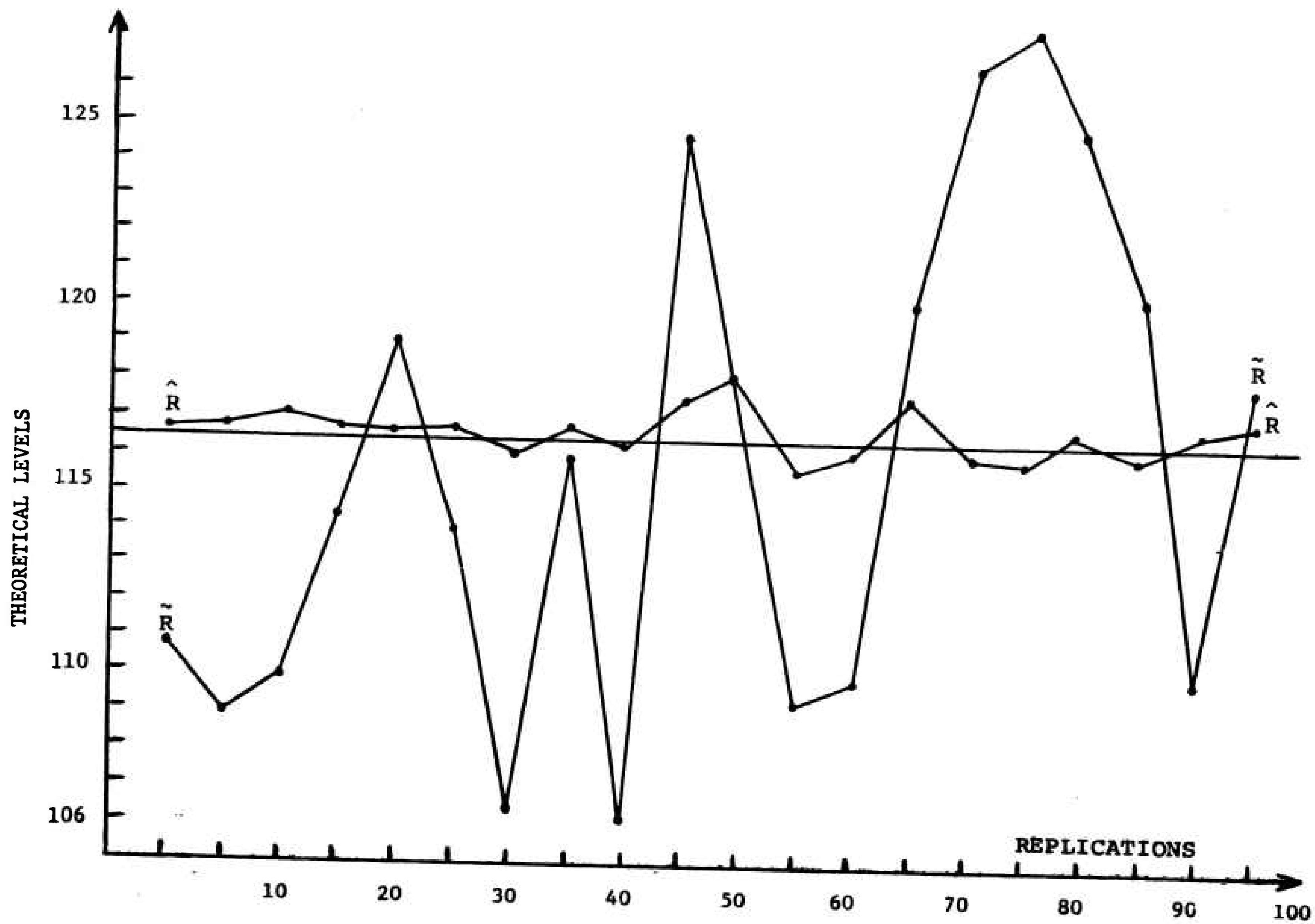

Figure 2.1. Comparison of $\tilde{R}$ and $\hat{R}$ for $\mu = 100$, $\sigma = 10$, $\rho = .05$.

and once again the strikingly larger variability in $\tilde{R}$ may be noted. One way to view these results is as follows. Think of 100 supply centers all operating under the same reorder rules for a given item. After 1,000 periods (far in excess of the number of periods for which records are typically kept) the graph may be viewed as showing the actual reorder levels that would be set at the various centers, first, all using $\hat{R}$ and, secondly, all using $\tilde{R}$. Again the results speak for themselves.

3. LINEAR MEAN MODEL.

As previously remarked, it may be unfair to indict exponential smoothing on the basis of a constant mean model since it appears to be designed more for models which are more time-dependent. Indeed, at the very heart of smoothing techniques is the idea that the most recent demands are more indicative of the true demand pattern than are the earlier ones. For a constant mean demand of course, that is not true and all demands reflect the true pattern equally well. But even when the mean is changing in time, this idea of weighting the most recent demand heavily must not be carried too far. For deterministic demands there can be little argument, but when demands are truly random, sudden increases or decreases in demand are to be expected even with a stable mean and there is a question of just how much weight should be assigned these random fluctuations. In any event the system can be studied to see what such effects are.

Perhaps the simplest time-dependent model that can be investigated is the case where demand is random with linear mean but a constant variance. This is the familiar linear regression model and with normality further assumed leads to standard maximum likelihood estimates of the parameters involved and once again presents itself as an alternative to exponential smoothing. To be more specific, suppose demand in period t is given by

$$(3.1) \qquad X_t = a + bt + \xi_t \quad \text{where} \quad \xi_t \quad \text{is} \quad N(0,\sigma^2)$$

Again Brown [3] recommends forecasting demands by means of exponentially smoothed estimates. This time, since two parameters are involved, a combination of single and double smoothing is required. More specifically, Brown advocates estimates

$$(3.2) \qquad \tilde{x}_t = 2\, S_t(x) - S_t^2(x)$$

$$\tilde{b} = \frac{\alpha}{\beta}\, [S_t(x) - S_t^2(x)]$$

Since $\mu_{t+1} = a + b(t+1) = \mu_t + b$, it follows that $\tilde{x}_t + \tilde{b}$ is a reasonable way of estimating μ_{t+1}. In these formulas, $S_t(x)$ stands for single smoothing applied to the demand record $X_0, X_1, X_2, \ldots, X_t$

$$S_t(x) = \alpha \sum_{k=0}^{t} \beta^k X_{t-k} \; ; \quad \beta = 1 - \alpha$$

$S_t^2(x)$ on the other hand represents smoothing applied to the sequence $S_0(x), S_1(x), S_2(x), \ldots, S_t(x)$ so that

$$S_t^2(x) = \alpha \sum_{k=0}^{t} \beta^k S_{t-k}(x)$$

It should be noted that $\hat{x}_t$ is not an estimate of a but rather of $a + bt$. With $\hat{b}$ given however, one can estimate a by the formula

$$\tilde{a} = \tilde{x}_t - \tilde{b}t \tag{3.3}$$

The reason for this observation is that usually in regression models of this type, estimates of the separate parameters are given. Indeed, in this notation, the maximum likelihood estimators of a and b are given by,

$$\hat{b} = \frac{\sum_{k=0}^{t} (k-\bar{k})(X_k-\bar{X})}{\sum_{k=0}^{t} (k-\bar{k})^2}, \qquad \bar{k} = \frac{1}{t+1} \sum_{i=0}^{t} i = \frac{t}{2} \tag{3.4}$$

$$\hat{a} = \bar{X} - b\bar{k}$$

There are standard formulas that may be found in almost any standard textbook on the subject. In these terms, an estimate of $\mu_{t+1} = a + b(t+1)$ would be given by $\hat{a} + \hat{b}(t+1) = \hat{a} + \hat{b}t + \hat{b}$.

This leaves the unknown parameter σ to estimate. In the theory of maximum likelihood, this estimate is easily derived and is given by considering average squared deviations about the fitted regression line. We thus have,

(3.5) $$\hat{\sigma} = \sqrt{\frac{1}{t+1}\sum_{k=0}^{t}(X_k-\hat{a}-\hat{b}k)^2}$$

Not surprisingly (in terms of Section 2) the parameter σ is estimated in exponential smoothing by looking at weighted absolute deviations about the fitted line. Thus we first let

$$e_t = X_t - \tilde{X}_{t-1} - \tilde{b}$$

be the difference between what was observed and what was forecast and then define

$$\tilde{\Delta}_t = \alpha \sum_{k=0}^{t-1} \beta^k \, |e_{t-k}|$$

With normality assumed we may then use

$$\tilde{\sigma} = \frac{\sqrt{\pi(2-\alpha)}}{2}\tilde{\Delta}_t$$

as before to estimate σ.

Once we have estimates of the various parameters of course we may use these to set reorder levels once again. In the spirit of the preceding section, two methods will be compared again. First, maximum likelihood estimates are used in each period to define

(3.6) $$\hat{R} = \hat{a} + \hat{b}(t+1) + k\,\hat{\sigma}$$

This would be the reorder level set at time t based on the fact that the "best" estimate of the next demand would be $\hat{a} + \hat{b}(t+1)$ the

estimate of the mean μ_{t+1}. If exponential smoothing is employed, then the reorder level would be set at

$$\tilde{R} = \tilde{x}_t + \tilde{b} + k\,\tilde{\sigma} \tag{3.7}$$

based on the same kind of reasoning.

How do these two methods compare? Again, we were forced to resort to simulation for reasons that are even more pronounced in this case. Generally speaking, and not too surprising perhaps, the two methods compared quite favorably with each other when attained risks were examined. The variability in the smoothing technique was not nearly so noticeable as it was in the constant mean case. Nevertheless, it was still present and never was reduced to an extent where it could be labeled superior to maximum likelihood in any of the cases examined.

First of all, many different cases (choices of a, b and σ) were examined by Coventry. For each parameter choice, estimates of the parameters were calculated by both methods after 100 periods of demand generated to satisfy the model of (3.1). The experiment was then replicated 100 times and results were then averaged over these cases, it was noted that the results appeared to be independent of parameter choices and so attention was focused on just a few special cases.

Typical of the results are those shown in Table 3.1 for the choice $a = 50$, $b = 2$ and $\sigma = 5$. The attained risks are displayed

for various periods and for this case averaged over 1,000 replications of the experiment. While the attained risks, $\tilde{\rho}$ using smoothing and $\hat{\rho}$ using maximum likelihood, are both reasonable close to the theoretical risk ρ, it should be noted once again that $\tilde{\rho}$ does tend to be more variable with no consistent pattern of change. In nearly every case $\tilde{\rho}$ does exceed $\hat{\rho}$ however and that in itself is noteworthy.

Number of Periods

	10		20		50		100	
$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$	$\tilde{\rho}$	$\hat{\rho}$
0.01	.036	.007	.027	.014	.024	.011	.029	.008
0.05	.103	.045	.070	.045	.078	.049	.068	.053
0.10	.155	.092	.124	.098	.137	.103	.114	.110
0.25	.306	.233	.264	.256	.274	.249	.249	.256
0.50	.516	.482	.519	.521	.526	.510	.517	.505

Table 3.1. Attained Risks Compared to Theoretical Risk

With these results and the many other cases examined, it is now reasonably safe to conclude that exponential smoothing is not a superior estimating technique for normal demands whether constant or linear in time. At least this is so when stock-out risk is the major criterion (as it often is) and when the methods presently employed by NavSup as advocated by Brown for setting reorder levels are compared to classical techniques. Indeed, depending on the consequences of facing an attained risk that is not the intended one, this method may be inferior to ordinary maximum likelihood techniques. This just about leaves computing

ease as the only criterion offered by smoothing advocates of any merit. But we found no evidence in any of our tests that smoothing resulted in any significant savings in computer time either. In most cases, the difference, if measurable, was negligible.

4. A BAYES PROCEDURE.

In actual practice it was found that neither the constant mean model nor that of the linear mean adequately reflects the true nature of demand even when the assumption of normality is acceptable. The model that comes closest to reflecting what most people involved really believe in (at least for some items) is that demand is normal with constant mean for a time, perhaps several periods, and then shifts to a new mean level which again remains constant for a time. For example, in times of conflict there may be a sudden increase in demand for an item and that demand has a mean value that remains fairly constant for the duration. But, as hostilities cease, the mean demand drops to a lower level and remains there while the circumstances remain stable. Then neither of the preceding models apply exactly although, subject to the general remarks previously made, exponential smoothing should be a good candidate for such a model. The reason is the often quoted property of responding to changes in demand more quickly than classical methods.

There is yet another technique which would seem quite appropriate for a model of this type and that is to apply Bayesian methods

sequentially to predict or forecast demand. The basic idea is to use posterior information in each period as prior information for the next period. Starting with some initial subjective judgment as to the parameters involved in the model, one can then proceed to use the information in each period to update one's guess as to the parameters to come and forecast accordingly. After all, if parameters such as mean demand are truly changing, possibly from period to period, then this basic Bayesian approach is tailored to fit precisely that kind of situation.

To be more specific, let us suppose that demand is still normally distributed but the mean is changing possibly from period to period. Initially, we also assume that the variance σ^2 in the initial period is known. In each period we will set a reorder level at a value $\mu^* + k\,\sigma^*$ where μ^* and σ^* are estimates of μ and σ for that period with k selected again in order to achieve some nominal risk ρ. To put these assumptions into the Bayesian framework, we initially assume that the conditional distribution of demand X given a value of the mean μ is normal with that mean and a known variance of σ^2. As to the mean μ, we suppose that the prior distribution on μ is normal with some mean μ_0 and variance σ_0^2. With this kind of normal on normal assumption, it is easy to show (see [6] for example) that the posterior distribution for μ, given an observed demand x_1, is again normal with mean μ_1 and variance σ_1^2 given by,

$$\mu_1 = \frac{\sigma_0 x_1 + \sigma^2 \mu_0}{\sigma_0^2 + \sigma^2}$$

(4.1)

$$\sigma_1^2 = \frac{\sigma_0^2\,\sigma^2}{\sigma_0^2 + \sigma^2}$$

As a bonus, if we take loss to be squared error, then the mean μ_1 of this posterior distribution is the Bayes estimator, meaning that it minimize the Bayes risk for the problem. (See [6] again for details.) As such, μ_1 and σ_1 are the best estimates--best from a Bayesian point of view--of the parameters that exist in nature at that point, namely, after one observation. Consequently, a logical Bayesian reorder level would be set at $\mu_1 + k\,\sigma_1$ and the corresponding stock-out risk $\rho_1 = P(X>\mu_1+k\sigma_1)$ may or may not be the target value ρ depending on whether or not the mean and standard deviation of demand in the second period are or are not μ_1 and σ_1 respectively.

Before proceeding to the next period it might pay to pause and analyze the significance of the estimates in (4.1). Re-writing μ_1 as

$$\mu_1 = \left(\frac{\sigma_0^2}{\sigma_0^2+\sigma^2}\right)x_1 + \left(\frac{\sigma^2}{\sigma_0^2+\sigma^2}\right)\mu_0$$

we see that the updated estimate of the mean based on the first observed demand x_1 is just a weighted average of x_1 and μ_0, the initial estimate of the mean. In this way, the weight attached to the observation x_1 is like a smoothing constant and may be used to reflect one's

desires or beliefs in the initial states. By choosing σ_0^2 small, very little (relatively) weight is attached to x_1 compared to μ_0. This is as it should be for if σ_0^2 is small, then the prior distribution is concentrated heavily about its mean μ_0 and reflects a high degree of credence in that initial choice μ_0. On the other hand, if one's initial belief in μ_0 is somewhat weak, this can be reflected by making σ_0^2 relatively large, whence more relative weight is attached to what is actually observed in x_1.

Fortunately, this same basic scheme continues from period to period as follows. By taking the prior for the mean in period 2 to be the posterior from period one, the posterior distribution for period 2, based on observing x_2, the actual demand during that period, is again normal with mean μ_2 and variance σ_2^2 given by the formulas,

$$\mu_2 = \frac{\sigma_0^2(x_1+x_2) + \sigma^2\mu_0}{2\sigma_0^2 + \sigma^2}, \qquad \sigma_2^2 = \frac{\sigma_0^2\,\sigma^2}{2\sigma_0^2 + \sigma^2}$$

Proceeding by induction in this manner, it is easy to show that the posterior distribution at the end of the period t based on having observed $x_1, x_2, \ldots, x_t$ is once again normal with mean μ_t and variance σ_t^2 given by

$$\mu_t = \frac{\sigma_0^2 \sum_{k=1}^{t} x_k + \sigma^2\mu_0}{t\sigma_0^2 + \sigma^2}$$

(4.2)

$$\sigma_t^2 = \frac{\sigma_0^2\,\sigma^2}{t\sigma_0^2 + \sigma^2}$$

As previously remarked, the reorder level for this period is then set at $\mu_t + k\,\sigma_t$.

Once again it may be seen that the Bayes estimate of mean demand, as given by the mean of the posterior distribution, is a weighted sum. This time the total observed demand $\sum_{k=1}^{t} x_k$ is weighted against the initial estimate μ_0. It is significant to note, moreover, that $\lim_{t\to\infty} \sigma_t^2 = 0$ so that, as time goes on, this posterior distribution is becoming degenerate at μ_t. Consequently, as a prediction of mean demand, the chosen value is subject to less and less fluctuation as time goes on.

So much for theory. To determine just how much the estimate of mean demand is affected by various combinations of μ_0 and σ_0^2 and to see how it compares with exponential smoothing, a pilot study using simulation was conducted. No attempt was made at this point to examine the behavior for the case of a shifting mean. Rather this study was confined to testing the procedure for internal consistency. For the case of a constant mean value of $\mu = 100$ and a choice of $\sigma = 10$, random demand was generated for 100 periods. The Bayes estimate was then computed for various a priori combinations of $\mu_0 = 0,\mu,\mu/2,\mu/3,\mu/4,\mu/5$ and $\sigma_0^2 = \sigma^2,\sigma^2/2,\sigma^2/3,\sigma^2/4,\sigma^2/5,2\sigma^2,3\sigma^2,4\sigma^2,5\sigma^2$.

The results are displayed in Table 4.1 the entries being the Bayes estimates or posterior means after 5 and 100 periods of observation. Obviously, the closer that μ_0 is to the true value of μ (100 in this case) the better the resulting estimate is. For large

σ_0^2 \ μ_0	0		20		25		33		50		100	
	5	100	5	100	5	100	5	100	5	100	5	100
20	50.2	95.3	60.2	96.2	62.7	96.5	66.9	96.9	75.2	97.2	100.2	100.0
25	55.8	96.2	64.7	97.0	66.9	97.2	70.6	97.5	89.0	98.1	100.2	100.0
33	62.8	97.1	70.3	97.7	72.1	97.9	75.3	98.1	81.5	98.6	100.3	100.0
50	71.7	98.1	77.4	98.5	78.9	98.6	81.3	98.7	86.0	99.1	100.3	100.0
100	83.7	99.1	87.0	99.3	87.8	99.3	89.2	99.4	92.0	99.5	100.3	100.0
200	91.3	99.5	93.1	99.6	93.6	99.7	94.3	99.7	95.8	99.8	100.4	100.0
300	94.1	99.7	95.4	99.8	95.7	99.8	96.2	99.8	97.3	99.9	100.4	100.0
400	95.6	99.8	96.6	99.8	96.8	99.9	97.2	99.9	98.0	99.9	100.4	100.0
500	96.6	99.8	97.3	99.9	97.5	99.9	97.8	99.9	98.5	99.9	100.4	100.0

Table 4.1. Bayes Estimates of Mean Demand (μ = 100, σ = 10)

values of the ratio σ_0^2/σ^2, it should be noted that the convergence to 100 is fairly rapid even for poor initial guesses. For example, with $\mu_0 = 0$ but $\sigma_0^2 = 5\sigma^2$, $\mu_5 = 96.6$ even after only 5 periods of observation.

Having thus tested the Bayes technique for internal stability, simulations were further used to compare the technique with exponential smoothing. For this comparison, mean demand was estimated by smoothing techniques using the formula,

$$\tilde{\mu} = S_t(x) = \alpha \sum_{k=0}^{t} \beta^k X_{t-k}$$

allowing for initial conditions $S_0(x)$ other than zero. Once again parameter choices $\mu = 100$ and $\sigma = 10$ were adopted. As a first comparison, the least favorable initial conditions, $\mu_0 = 0$ and $S_0(x) = 0$, were selected. Estimates of mean demand over various periods were then made for a variety of choices of the weighting factor σ_0^2 and the smoothing constant α. The results are reported in Table 4.2 where it may be seen that Bayes estimates are typically better than those given by exponential smoothing when roughly the same relative weight is attached to the observations. Thus, small values of σ_0^2 should be compared with small values of α. If we take $\alpha = 0.2$ as presently used by NavSup as a guide, then almost any choice of σ_0^2 will do better in the early stages and about as well in later periods.

$\mu = 100$ $\sigma = 10$	$\mu_0 = 0$ $S_0(x) = 0$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_0^2 = \sigma^2$	83.7	91.0	93.7	97.9	99.1
	$2\sigma^2$	91.3	95.4	96.8	98.8	99.5
	$3\sigma^2$	94.1	96.9	97.8	99.2	99.7
	$4\sigma^2$	95.6	97.7	98.3	99.3	99.8
	$5\sigma^2$	96.6	98.2	98.7	99.4	99.8
SMOOTHING	$\alpha = 0.1$	41.1	65.1	79.3	99.2	100.4
	0.2	67.3	89.1	96.4	99.8	100.6
	0.3	83.2	96.9	99.5	99.9	100.8
	0.4	92.1	99.0	100.0	99.9	100.9
	0.5	96.5	99.5	100.3	99.9	100.9

Table 4.2. Estimates of $\mu = 100$

To assess the effect of initial conditions or risk, the same basic model was used to generate demands for 100 periods. Reorder levels were then set on the basis of a k-value to achieve a theoretical risk of $\rho = .05$ using both techniques and the actual attained risk was then recorded. The experiment was then replicated 1,000 times and the attained risks averaged over these replications. The results are reported in Table 4.3 for the worst initial conditions $\mu_0 = 0 = S_0(x)$ and in Table 4.4 for the best initial conditions $\mu_0 = 100 = S_0(x)$.

The results are quite remarkable. Except for a few cases the Bayes method provides a sample risk closer to the theoretical one than does exponential smoothing even for poor initial conditions. In both cases, when a small value of α is chosen the long term results are fairly accurate, but the results in the early periods are far from satisfactory. For large values of α the results in the early periods are better but only at the expense of weaker results in later periods. By comparison, the Bayes technique produces about the same results in any case. For large weighting constants $(\sigma_0^2 = 5\sigma^2)$, the Bayes method adjusts quite rapidly and the long term results are all fairly accurate.

Of course, all of these results are average values, averaged over the replications. How badly they vary from one replication to another is important also. To check on variability, the sample standard deviations of the estimates of μ for the 1,000 replications were computed. Those values are reported in Table 4.5 for the case

$\mu = 100$ $\sigma = 10$	$\mu_0 = 0$ $S_0(x) = 0$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_0^2 = \sigma^2$	.634	.294	.187	.079	.071
	$2\sigma^2$	.310	.159	.107	.068	.062
	$3\sigma^2$	.211	.128	.090	.066	.061
	$4\sigma^2$	.168	.109	.084	.065	.061
	$5\sigma^2$	.146	.097	.080	.065	.060
SMOOTHING	$\alpha = 0.1$	1.000	.992	.737	.067	.063
	0.2	.995	.390	.145	.061	.069
	0.3	.761	.137	.080	.070	.079
	0.4	.365	.088	.080	.076	.083
	0.5	.187	.084	.083	.084	.091

Table 4.3. Sample Risks with Worst Initial Conditions

$\mu = 100$ $\alpha = 10$	$\mu_0 = 100$ $S_0(x) = 100$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_0^2 = \sigma^2$	.066	.056	.057	.057	.055
	$2\sigma^2$	.069	.057	.058	.057	.055
	$3\sigma^2$	.072	.058	.058	.057	.055
	$4\sigma^2$	.073	.059	.058	.057	.055
	$5\sigma^2$	.074	.059	.058	.057	.055
SMOOTHING	$\alpha = 0.1$	.060	.051	.057	.057	.063
	0.2	.059	.061	.061	.061	.069
	0.3	.066	.073	.069	.070	.079
	0.4	.073	.077	.079	.076	.083
	0.5	.082	.082	.083	.084	.091

Table 4.4. Sample Risks with Best Initial Conditions

$\mu = 100$ $\sigma = 10$	$\mu_0 = 100$ $S_0(x) = 100$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_0^2 = \sigma^2$	3.95	3.26	2.84	1.95	1.69
	$2\sigma^2$	4.25	3.40	2.94	1.97	1.72
	$3\sigma^2$	4.36	3.41	2.94	1.96	1.73
	$4\sigma^2$	4.46	3.47	2.95	1.95	1.72
	$5\sigma^2$	4.48	3.48	2.97	1.95	1.72
SMOOTHING	$\alpha = 0.1$	2.31	2.64	2.68	2.61	2.73
	0.2	3.41	3.68	3.63	3.60	3.65
	0.3	4.30	4.51	4.40	4.49	4.44
	0.4	5.10	5.27	5.17	5.30	5.16
	0.5	5.96	6.03	5.90	6.06	5.88

Table 4.5. Sample Standard Deviations

$\mu_0 = 100 = S_0(x)$. It may be seen that for early periods and small choices of α, the smoothing method is less variable. But, as more and more periods are taken, the Bayes method produces less variable results, indeed the standard deviation consistently decreases with time. On the other hand, smoothing yields results that appear to have about the same variance regardless of how many periods are observed, a phenomenon that has been noted before. The sample standard deviations for the case $\mu_0 = 0 = S_0(x)$ were surprisingly about the same as the most favorable case and are not presented here.

5. CONCLUDING REMARKS

Regardless of what else might be said about exponential smoothing as a forecasting tool, it now seems reasonably safe to say that the results tend to be more variable than some other alternative methods that are available. This same basic theme keeps recurring in model after model and case after case. Claims in this regard have repeatedly been made with due caution throughout this and earlier studies due to the simulation techniques employed. Yet the consistency of recurrence, coupled with the large sample sizes used, cannot be safely ignored. In some isolated cases, we have supplied a theoretical basis for the observations.

It is practically never the case that exponential smoothing dominates the alternatives studied regardless of the criterion used for comparison. One of the outgrowths of this study is to highlight

the importance of variance whenever random demand is faced. It is a quantity that must be reckoned with, for it is of little comfort to the individual inventory manager to know that his technique does well on the average unless some idea of the variability is also known.

Of the alternatives studied, the Bayes method of Section 4 seems admirably suited to a model where mean demand is constant in a given period but subject to change from period to period. The method supplies a natural and appealing method of incorporating information on a prior basis to update estimates sequentially as information is gathered. More needs to be done with the method, however, before it can be endorsed over other alternatives. This would be the basic recommendation of this study, which should be viewed only as an initial pilot study of this technique. Another recommendation would be to urge all users of exponential smoothing to give serious consideration to testing other alternatives in the particular context of their special application. Special attention should be paid to at least replacing MAD as a method of estimating variance. This much change alone may produce less variable results and thereby make a stronger case for exponential smoothing.

BIBLIOGRAPHY

[1] Zehna, Peter W., Some Remarks on Exponential Smoothing, Naval Postgraduate School, Technical Report No. 72, December 1966.

[2] Zehna, Peter W., Forecasting Errors Using MAD, Naval Postgraduate School, Technical Report NPS55Ze9041A, April 1969.

[3] Brown, R. G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall, Inc., 1963.

[4] Ornek, O., An Analysis of Mean Absolute Deviation Variability, M. S. Thesis, Naval Postgraduate School, October 1969.

[5] Coventry, J. A., A Comparison of Demand Forecasting Techniques, M. S. Thesis, Naval Postgraduate School, March 1971.

[6] De Groot, M. H., *Optimal Statistical Decisions*, McGraw-Hill Book Company, 1970.